ELSEVIER

Learning to see faces and objects

Michael J. Tarr and Yi D. Cheng

Cognitive and Linguistic Sciences, Brown University, Box 1978, Providence, RI 02912, USA

Visual recognition of objects is an impressively difficult problem that biological systems solve effortlessly. We consider two aspects of this ability. First, is the recognition of all objects accomplished by either a single system or multiple, domain-specific systems? Behavioral, neuropsychological and neuroimaging data indicate that a single system is sufficient for the recognition of all objects at all levels. Second, how does such a system 'tune' itself to the constraints imposed by recognition at different levels of specificity? Evidence indicates that the task demands and learning that arise from different forms of feedback determine which computational routines are recruited automatically in object recognition.

The truck driver slammed on the brakes, not expecting the road to be full of monkeys.

[Reported during a lecture by our colleague, Philip Lieberman.]

Rarely does one expect to see a road full of monkeys. Yet these objects, although surprising when out of context, are easily interpreted. Such is the remarkable nature of object perception by humans, within seconds of viewing an unexpected scene, that we know it contains many animals, that they are monkeys and, possibly, that they are macaques and that one is named Curious George. Our object-perception abilities are noteworthy not only for their speed and accuracy, but for their flexibility – we recognize objects at multiple levels of specificity in the absence of any prior expectations. How the brain solves this problem has been a subject of increasing interest for the past 20 years.

At the heart of the problem are two questions. First, what processes are recruited across recognition tasks? One possibility is that a single system is sufficient to recognize all objects at all levels of specificity. Another possibility is that multiple domain-specific systems are deployed, depending on object category and task demands. Second, what are the origins of our object-recognition abilities? Are they innate or acquired through experience? To answer these questions we begin by considering two dimensions along which systems for object recognition have been characterized.

Recognition of different object categories

One taxonomy for object recognition relies on the fact that objects can be placed into distinct categories based on their visual appearance [1,2]. Objects that are perceptually similar, that share similar shape or surface properties, can be grouped together at either the basic-level [1] or, depending on experience, at the entry-level [2]. For example, a primatologist would distinguish between vervets and macaques, whereas most of us would treat both as monkeys. Given the existence of distinct visual categories, many researchers have posited the existence of multiple recognition systems, each of which support the recognition of one or more visual categories. The most popular distinction has been that there are separate systems for faces and non-face objects [3–8]. This dichotomy is based on intuitions such as the inherent social significance of faces and the difficulty of discriminating individual faces, as well as the following evidence:

(1) A visual preference for face-like stimuli in neonates [9,10]
(2) Face-specific effects in behavioral measures of visual processing [11,12]
(3) Face-selective neurons [13], brain areas [5,14] and neural signals [6]
(4) Dissociations between face and object recognition in brain-injured patients [3,4,7,15]

Although these points make a seemingly convincing case for separate systems for faces and objects, they are based on the questionable assumption that there are characteristics and modes of processing that are exclusive to faces and face recognition. Alternatively, these properties may be ones that faces happen to have, but they could also be true for other object categories. Thus, if the characteristics and computational mechanisms involved are not face specific, it is possible that a single system might support recognition of both face and non-face objects. Indeed, when factors other than the visual category are considered, such as the specificity of the recognition judgment and the degree of expertise with that category, faces and objects elicit similar patterns of behavioral and neural responses (Box 1).

Recognition at multiple levels

Although current evidence does not appear to support separate recognition systems for faces and objects, the analysis in Box 1 does indicate a second possible taxonomy for object recognition. Different behavioral and neural responses are observed when we recognize objects at more specific levels compared with the basic level (e.g. either 'Curious George' or 'macaque' as opposed to 'monkey') [16–18] (Fig. 1). Thus, many researchers have suggested that we have one recognition system for more specific-level judgments and another for basic-level judgments [19–22]. The crucial idea is that the former is accomplished using specific features of an image to discriminate between objects whereas the latter is accomplished by mapping

Corresponding author: Michael J. Tarr (michael_tarr@brown.edu).

http://tics.trends.com 1364-6613/02/$ - see front matter PII: S1364-6613(02)00010-4

Box 1. Is face recognition 'special'?

Although many studies in the literature cite behavioral or neural dissociations between face and object recognition as evidence for separate processing systems, apparent specializations in the processing of faces may be attributed to the default level at which faces are recognized – individual identity. Putatively 'face-specific' effects might actually be elicited by factors that are not exclusive to faces. For example, recognizing objects at a more specific level recruits the same neural substrates that are recruited by face recognition [a,b]. This indicates that perceptual expertise leads to a shift in the default level of recognition within an expert's domain [c]. Thus, birders cannot help but recognize individual species of birds and dog-show judges cannot help but recognize individuals within a breed [c]. In this context, face recognition may be considered a case of perceptual expertise acquired by almost everyone [d]. With the onset of expertise the processing necessary for more specific-level categorization becomes automatic. These processes are recruited when objects are recognized intentionally at a more specific level [a,b] or in a domain of expertise. Thus, the dissociations between faces and objects observed in behavior, neuropsychology and neuroimaging might be reinterpreted as dissociations in the level of recognition and the degree of expertise between two object categories.

Evidence to support this interpretation comes from behavioral studies in which dog-show judges sometimes exhibit a similar inversion effect for recognizing faces and dogs [e]. Moreover, subjects trained to be experts at recognizing Greebles (see Fig. 2, main text) often exhibit similar configural effects when recognizing faces and Greebles [f–h]. These behavioral commonalities are reflected in the neural coding of faces and objects because comparable patterns of focal neural activity are obtained in the, so-called, fusiform face area (FFA) of Greeble experts during recognition of either faces or Greebles [i]. Likewise, the FFA area of bird or car experts is activated when recognizing birds or cars, respectively [j] (see Fig. I). Recordings of event-related potentials reveal that bird and dog experts have an enhanced N170 occipitotemporal component for their domain of expertise relative to non-expert domains [k]. Similarly, Greeble experts (but not novices) exhibit a delay and enhancement of the N170 when recognizing inverted faces or Greebles (left lateralized for Greebles) [l]. Last, prosopagnosics patients who have impaired ability to recognize faces have similar deficits in subordinate-level recognition of everyday objects and Greebles [m] (but see [n]). These data provide evidence that the same recognition system is recruited when recognizing either faces or objects at more specific levels, often automatized by expertise.

References

a Gauthier, I. *et al.* (1997) Levels of categorization in visual recognition studied with functional Magnetic Resonance Imaging. *Curr. Biol.* 7, 645–651

b Gauthier, I. *et al.* (2000) Does visual subordinate-level categorisation engage the functionally defined fusiform face Area? *Cogn. Neuropsychol.* 17, 143–163

c Tanaka, J.W. and Taylor, M. (1991) Object categories and expertise: is the basic level in the eye of the beholder? *Cogn. Psychol.* 23, 457–482

d Tanaka, J.W. (2001) The entry point of face recognition: evidence for face expertise. *J. Exp. Psychol. Gen.* 130, 534–543

e Diamond, R. and Carey, S. (1986) Why faces are and are not special: an effect of expertise. *J. Exp. Psychol. Gen.* 115, 107–117

f Gauthier, I. and Tarr, M.J. (1997) Becoming a Greeble expert: exploring the face recognition mechanism. *Vision Res.* 37, 1673–1682

g Gauthier, I. *et al.* (1998) Training Greeble experts: a framework for studying expert object recognition processes. *Vision Res.* 38, 2401–2428

h Gauthier, I. and Tarr, M.J. (2002) Unraveling mechanisms for expert object recognition: bridging brain activity and behavior. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 431–446

i Gauthier, I. *et al.* (1999) Activation of the middle fusiform face area increases with expertise in recognizing novel objects. *Nat. Neurosci.* 2, 568–573

j Gauthier, I. *et al.* (2000) Expertise for cars and birds recruits brain areas involved in face recognition. *Nat. Neurosci.* 3, 191–197

k Tanaka, J.W. and Curran, T. (2001) A neural basis for expert object recognition. *Psychol. Sci.* 12, 43–47

l Rossion, B. *et al.* (2002) Expertise training with novel objects leads to left-lateralized face-like electrophysiological responses. *Psychol. Sci.* 13, 250–257

m Gauthier, I. *et al.* (1999) Can face recognition really be dissociated from object recognition? *J. Cogn. Neurosci.* 11, 349–370

n Moscovitch, M. *et al.* (1997) What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *J. Cogn. Neurosci.* 9, 555–604

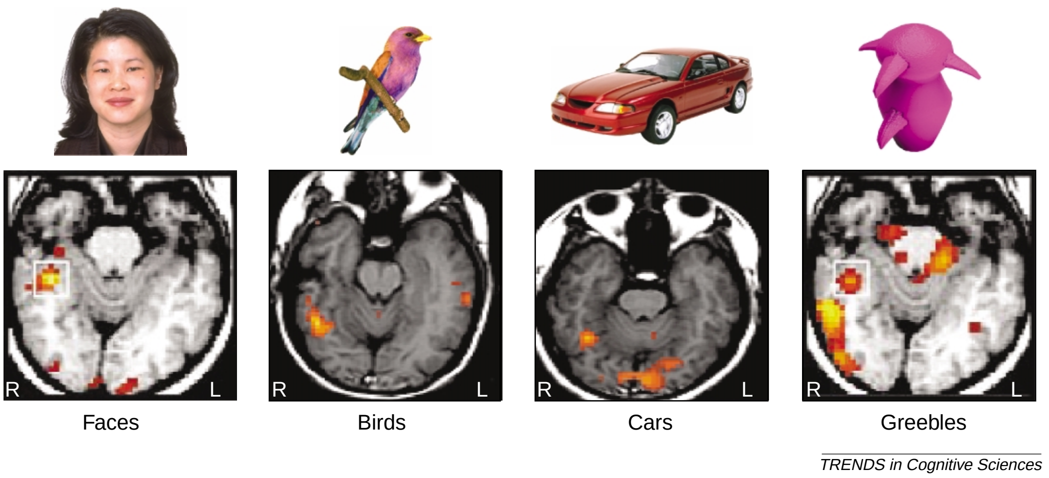


Fig. I. Examples of individual fMRI scans of experts for four different object categories. Each fMRI image shows the activation associated with viewing items from the depicted object category that was significantly *over and above* the activation associated with viewing common objects. The slice angles and activation thresholds for each image are not equated because these data were collected across different studies. However, in each case there is a significant cluster of activation in the right fusiform gyrus – sometimes referred to as the 'Fusiform Face Area' or 'FFA'. These studies demonstrate that visual expertise produces a pattern of neural activity for non-face objects in the domain of expertise that is quite similar to that obtained for faces. fMRI scans adapted from [47] and [50].

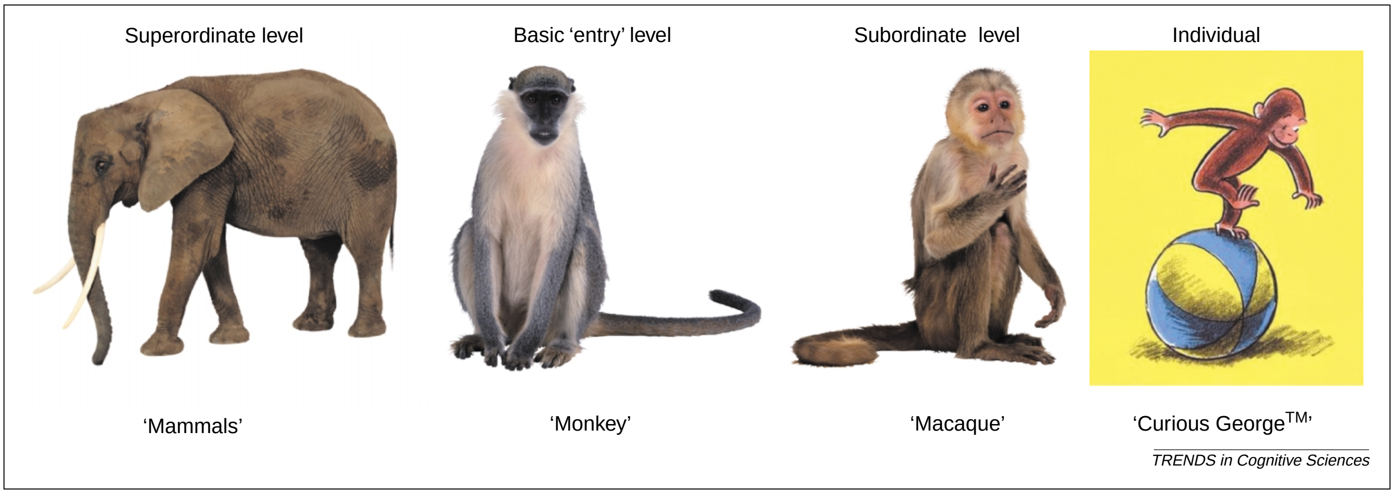


Fig. 1. Multiple levels of description. We are able to recognize objects at different levels of specificity depending on experience and task demands. For example, macaques and vervet monkeys would share similar visual descriptions at the basic (entry) level, but different descriptions at the species or subordinate level. Particular monkeys, such as Curious George™, are recognized at the individual level.

individual objects that are similarly shaped into a single representation, a visual category. Accomplishing a many-to-one mapping requires that the individual shape and surface details needed for specific-level recognition be filtered out or ignored, typically by using a coarse description of object shape provided by, for example, three-dimensional parts [19], skeletal models [23] and qualitative features [24]. Thus, macaques and vervet monkeys would share similar visual descriptions at the basic level, but different descriptions at the species level. This example illustrates that a crucial variable for basic-level recognition is the shape of the object [1]. However, if two objects that are nominally members of the same basic-level category have very different shapes they will be categorized visually as different objects. For example, although sparrows and penguins are both birds at the basic level, penguins are first recognized as penguins, not birds, because of their untypical bird shape. This distinction motivates the idea of the entry level [2], a term that is meant to capture the default level of visual categorization of everyday objects.

How do we evaluate whether the ability to recognize objects at multiple levels is supported by one or more recognition systems? One approach is to identify behavioral markers for each putative system and test whether these patterns of behavior are consistent across different levels of recognition. For example, putative basic-level and specific-level systems are often distinguished by their degree of viewpoint sensitivity [18–20,22]. Perhaps the strongest prediction is that basic-level recognition will be invariant across many transformations of the object. More specifically, changing the appearance of an object by, for example, rotation, translation and altering the lighting will not change the description of the object as long as the post-change image presents the features seen previously [19,25,26]. This analysis is consistent with the many models of entry-level recognition that posit equivalent recognition speed and accuracy over either three-dimensional rotation of an object or observer movement in three-dimensions, but slowed and less accurate performance for specific-level categorization [18–20,22]. Alternatively, models that assume a single system for recognition at all levels of specificity posit that recognition speed and accuracy vary with the difficulty of a task but remain viewpoint-dependent when there is some change in the appearance of the object [27–29].

Testing these alternatives has been problematic because there is little consensus regarding appropriate stimuli and tasks [30,31]. What emerges is that viewpoint dependence or invariance is not a dichotomous dimension – indeed, every time one recognizes an object from even a slightly changed viewpoint there is some cost in performance (Box 2). More importantly, the magnitude of this cost seems to be modulated by the specificity of the recognition judgment in a continuous manner; the more similar an object is to potential incorrect targets, the higher the cost for changes in viewpoint [27–29]. Consequently, it is difficult to say where the boundary is between the two putative systems. Other apparent dissociations between basic-level and specific-level recognition also appear to be along a continuum and patterns of behavior that are diagnostic for one system hold for tasks associated with the other [32,33]. Thus, there is little evidence to indicate that systems for basic-level versus specific-level recognition can be separated. This is not to say that there are not significant differences in the manner in which basic-level and specific-level judgments are accomplished, but these differences might reflect a variety of perceptual processes in the toolkit available to a single, flexible system (Box 2) [29,34,35].

How do we learn to see objects and faces?

How does the visual recognition system 'tune' itself to use this toolkit of processes according to the constraints imposed by recognition at different levels of specificity? One possibility is that we are born prewired to handle the different demands implicit in learning and recognizing all objects at all levels [36]. This suggestion relies on innate mechanisms, which implies a modular system [5,8,37]. A less extreme stance posits that the process of object recognition interacts with our biases, intentions and experience in a nontrivial manner (i.e. more than raw

Box 2. Is object recognition ever viewpoint invariant?

In structural-description theories of object recognition, it is often hypothesized that object recognition is invariant over changes in either object orientation or observer viewpoint [a,b]. At the same time, view-based and image-based theories predict that object recognition is dependent on the appearance of the object as viewed originally [c,d]. Tests of these two alternatives vary, but the standard manipulation almost always presents an observer with an object from one view and measures their speed and accuracy when recognizing the same object from a new view. This basic paradigm consistently reveals systematic costs for changes in viewpoint of novel two-dimensional shapes [c], novel three-dimensional objects [e] and familiar, common objects rotated in either two [f] or three [g] dimensions (Fig. I). Not only does recognizing an object from a new view take longer and become less accurate, but the magnitude of this effect generally corresponds with the degree of viewpoint change (i.e. larger rotations produce proportionally poorer performance) [c,e–g].

Although observations of viewpoint-dependent recognition behavior are undeniable, there have been attempts to describe such findings as limited cases that do not extend to entry-level or 'everyday' recognition (presumed to be viewpoint invariant) [h]. However, common objects recognized at the entry level show viewpoint-dependent effects [f,g]. Moreover, the recognition of novel three-dimensional objects that are qualitatively different in shape (thereby mapping into different entry-level categories) is also viewpoint dependent [i,j]. All things considered, there are only a few instances where object recognition is viewpoint invariant. Such cases are a product of a particular concatenation of stimuli and tasks that bias observers to hone in on local, unique features that are diagnostic for individual identity [j]. Even then, this strategy is relatively unstable because including additional diagnostic features in each object prompts observers to shift to viewpoint-dependent processing [l]. Such results demonstrate that there is no 'hard' boundary between class-level and specific-level recognition mechanisms. Rather, there is a continuum of effects modulated by the level of recognition, the inherent similarity of the stimuli, and the task, which is perhaps best explained by a single recognition mechanism that is sensitive to these factors [k,l].

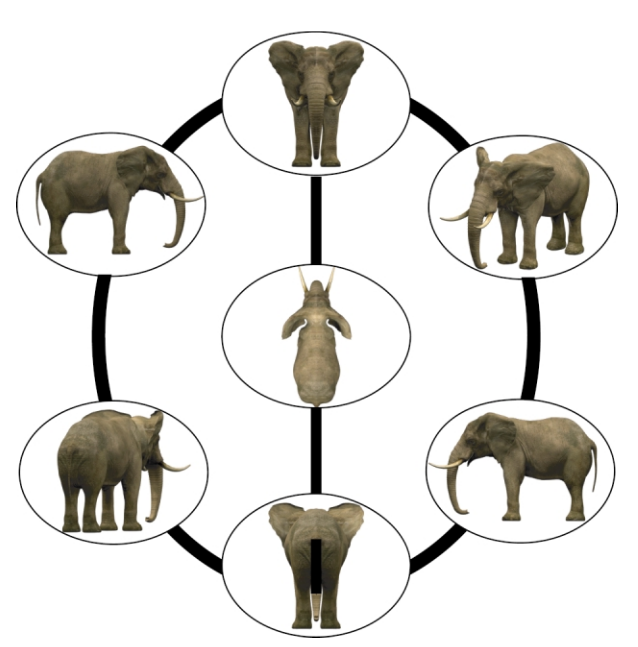


Fig. I. A viewpoint-dependent object representation of an elephant. Because individual views will only generalize over a limited range of positions, the representation must include multiple views of the same object, each showing different visible features.

References

a Marr, D. and Nishihara, H.K. (1978) Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. R. Soc. Lond. Ser. B* 200, 269–294

b Biederman, I. (1985) Human image understanding: recent research and a theory. *Comput. Vis. Graphics Image Process.* 32, 29–73

c Tarr, M.J. and Pinker, S. (1989) Mental rotation and orientation-dependence in shape recognition. *Cogn. Psychol.* 21, 233–282

d Poggio, T. and Edelman, S. (1990) A network that learns to recognize three-dimensional objects. *Nature* 343, 263–266

e Bülthoff, H.H. and Edelman, S. (1992) Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proc. Natl Acad. Sci. U.S.A.* 89, 60–64

f Jolicoeur, P. (1985) The time to name disoriented natural objects. *Mem. Cogn.* 13, 289–303

g Lawson, R. and Humphreys, G.W. (1996) View specificity in object processing: evidence from picture matching. *J. Exp. Psychol. Hum. Percept. Perform.* 22, 395–416

h Biederman, I. and Gerhardstein, P.C. (1993) Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance. *J. Exp. Psychol. Hum. Percept. Perform.* 19, 1162–1182

i Tarr, M.J. *et al.* (1997) To what extent do unique parts influence recognition across changes in viewpoint? *Psychol. Sci.* 8, 282–289

j Tarr, M.J. *et al.* (1998) Three-dimensional object recognition is viewpoint-dependent. *Nat. Neurosci.* 1, 275–277

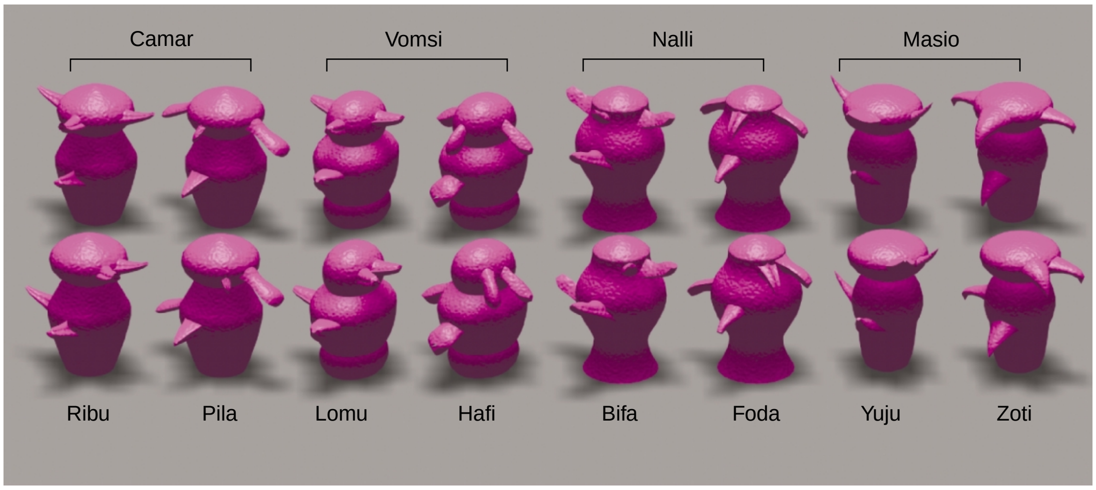
k Perrett, D.I. *et al.* (1998) Evidence accumulation in cell populations responsive to faces: an account of generalisation of recognition without mental transformations. *Cognition* 67, 111–145

l Tarr, M.J. (2003) Visual Object Recognition: Can a Single Mechanism Suffice? *Analytic and Holistic Processes in the Perception of Faces, Objects, and Scenes* (Peterson, M.A. and Rhodes, G., eds), Oxford University Press, (in press)

experience is required) [38,39]. For example, there are conditions under which semantic knowledge and task constraints affect recognition behavior [40]. In other words, we might recognize objects using more than just the information we extract from the scene (bottom-up processing). Specifically, top-down processes, such as the knowledge of an object's function, might play a role in determining which processes are deployed. Also, objects are not, at least by default, all recognized at the same level of visual granularity [35]. Typically, observers must learn to recognize objects at more subordinate levels of analysis. For example, our ability to recognize faces of a given race is partially a product of experience [41] and observers do not automatically distinguish between instances of a visual category without extensive practice [42–44]. Thus, throughout their lifetimes people constantly tune their visual-recognition abilities [35,45].

What can we learn?

All of us can group even the most oddly shaped novel objects into visual categories. At the same time, barring brain injury or disease, we all come to recognize faces by default at the individual level. Evidence indicates that, with interest from a hobby or a profession, we can also learn to discriminate instances of non-face objects at finer levels. For example, car buffs can discuss subtleties between different models of Porsches [46], birders can identify a myriad of wrens and finches [46,47], dog-show



TRENDS in Cognitive Sciences

Fig. 2. Examples of sixteen Greebles. The individuals in this novel, homogeneous visual category are organized into two 'genders' (left and right members of each Greeble pair) and five 'families' (four of which are depicted and identified by the top row of labels). Examples of individual Greeble names are given in the bottom row of labels. The top row illustrates symmetric Greebles, while the bottom row illustrates asymmetric Greebles.

judges can tell a prize Schnauzer from an also-ran [47,48], radiologists can diagnose a diseased lung from an x-ray and laboratory subjects can learn the individual names of 'Greebles' (see Fig. 2) [43,44].

Such examples indicate that visual expertise is not limited to faces and that we can learn to discriminate between individual items in a wide variety of visually homogeneous objects. Interestingly, the ability to acquire visual expertise might have evolved as a specific response to the social need to identify individual faces. Thus, although individual-level recognition is not exclusive to faces, it might still be optimized for face recognition in particular and, as such, be constrained in the types of object forms that can be individuated. This logic is inherent in the argument that Greebles (and dogs, cars, birds, etc.) elicit face-like behavioral and neural patterns in experts because they 'look like faces' [8,35]. More formally, the learning processes that support the acquisition of expertise might be tuned to certain properties that faces happen to have (but are not face-specific).

To the extent that researchers have 'pushed the envelope' by creating experts with novel stimuli [43,44,49] or testing experts with previously-learned stimuli [46–48], the visual recognition system has been revealed to be remarkably plastic. Accordingly, abilities and patterns of behavior similar to those observed for faces are often found for non-face objects. Moreover, we can identify the properties of faces that are also true for other known domains of expertise (i.e. define what is meant by 'looking like faces'). For example, one common property is that faces, dogs, cars, birds and Greebles have bilateral symmetry. However, only faces and Greebles have a 2–1–1 configuration of parts (eyes–nose–mouth), indicating that this configural property is not necessary. Similarly, although faces, dogs and birds are living things, cars and Greebles are neither living nor do they have surfaces that look organic. Thus, our recognition system appears to be capable of learning and recognizing a wide range of visual categories; although there may be some constraints on the geometries and images that can be learned, they appear to be very soft.

How do we learn?

Object recognition encompasses multiple processing levels, from the categorical to the specific. We have focused on the processes that support specific-level recognition, which might be the easiest problem solved by our visual-recognition systems. Pigeons, primates, people and many artificial neural-network models of recognition can quite easily learn specific images of objects. By contrast, generalizing across such images is much harder [25].

That specific images of objects are readily learned could be taken as evidence that expertise should not have to be acquired at all. Representations that support expert discriminations might be learned in the form of early visual responses (e.g. the output of sets of receptive fields [50]). Accordingly, the ability to recognize two images or configurations of an object as the same object, or recognize two objects as instances of the same category would need to be acquired. But, as noted above, this seems contrary to intuition. We are good at visually categorizing objects even at a young age [51] and, for the most part, entry-level recognition defaults to the basic level unless modulated by expertise [52]. It is acquiring this expertise that seems hard and, even for faces, takes years of experience [42]. Thus, although learning specific descriptions of images might be a straightforward consequence of early vision, this sort of representation does not translate into visual expertise. The question is why; what is the difference between learning specific images and discriminating between individual objects?

Consider that: (i) expert-level recognition for objects such as faces seems to be robust over noise [52], (ii) experts can learn new, individual instances rapidly, and (iii) in a domain of expertise, individual-level recognition is accurate and fast, even when discriminating between similar

Box 3. An analogy between language and vision

Our characterization of visual expertise is analogous to popular accounts of the acquisition of language [a]. For example, one classic problem was defined by Quine [b]: 'If one is walking with a native speaker of a language different from one's own and a rabbit runs across your path and your companion exclaims "Gavagai!", to what are they referring?' The 'poverty of the stimulus argument' suggests that you (and children learning a language) receive little feedback about the relationship between words and objects, yet you (and they) learn the correct mappings. One solution is that from birth, specific biases constrain our hypotheses about the relationship between words and objects [a]. Such biases, along with context, allow us to focus on the appropriate features of the environment and map words to these features without explicit feedback. In Quine's example, we may apply two biases: the unknown word is a count noun and that count nouns apply to object shapes [c]. Thus, 'Gavagai' refers to the object 'rabbit' because the learner assumes a label (particularly a single word) refers to an object's shape (and hence its identity) and not a substance or action. Similar constraints may be exerted in expertise acquisition; we may rely on innate and internally generated (e.g. hobbies) biases along with context to help bootstrap learning of new object domains.

References

a Pinker, S. (1994) *The Language Instinct*, William Morrow and Co
b Quine, W.V. (1960) *Word and Object*, MIT Press
c Landau, B. *et al.* (1988) The importance of shape in early lexical learning. *Cogn. Develop.* 3, 299–321

objects [53]. By contrast, object representations that are based on literal images do not generalize to either degraded or transformed versions. Furthermore, image-based similarity metrics that allow such generalizations would introduce new problems, such as confusion between visually similar exemplars in a visual category. Thus, expertise is different from simply learning images. Indeed, visual expertise is somewhat paradoxical because it seems to recruit specific representations for discriminating between individuals (indicated by the strong viewpoint, lighting and configural dependence of expert-level recognition [11,12,43,44,48]), but seems quite robust to transformations.

This characterization of expert-level recognition indicates that we learn object representations that are specific across some dimensions but flexible across others. In some ways this is not very different from the challenges faced by a learner at the basic level. For example, learning to associate two objects as members of the same category is similar to learning to associate two configurations of the same object at different moments in time. Crucially, the same learning mechanisms might apply: despite variations in the images, a high degree of visual similarity between the two targets relative to other known objects will lead to their association [22,54]. Likewise, because the same sets of features co-occur across multiple viewings, they will become more strongly associated with one another, thereby refining the representation at the configural level [22]. The operation of these and other associative mechanisms [55] are unsupervised, that is, they function without any feedback about the correctness or validity of the learning that has occurred. However, the acquisition of expertise requires more than raw experience – without explicit training or feedback it is unclear whether one could ever become an 'expert' in any domain.

What sort of feedback is needed to become an expert?

No feedback

Can one become a visual expert without feedback? Self-organizing neural-networks are capable of learning large numbers of patterns and the relationships between them without feedback or supervision [56]. Such pattern associators are sensitive to the statistical relations within the dataset. Thus, they can learn to discriminate individual images of objects and, to some extent, to generalize over different images of the same object or category. What is unknown is the extent to which self-organizing systems can account for expert-level recognition. Self-organizing object 'spaces' form a critical core for further learning [54]. Moreover, intuitively, some sort of supervision or intentional learning is essential to becoming an expert; although we see hundreds of cars everyday, few of us become car experts unless we expend the effort and interest needed to acquire expertise [35,43–45,49].

Implicit feedback

Where does this interest come from? In the case of faces it might be innate because there is evidence that newborns prefer to look at facial configurations (eyes, nose, mouth) over other patterns [10]. It is as if their brains are telling them to look at (and learn) faces. This preference for face geometry may be construed as a bias that helps to shape category learning by directing attention. Learning based on such biases might be dubbed 'internally supervised' in that although there is no explicit training signal, the system is directed to attend to some objects at the expense of others. This idea is consistent with a growing interest in feedback loops in cortex that seem critical for learning [57].

A second form of implicit feedback is the context in which an object appears. Expectations might help determine the accuracy of our recognition judgments. For example, we may fail to recognize a friend in an unexpected setting, whereas an expected context might facilitate identification. At a minimum, the acquisition of expertise relies on these and other forms of implicit feedback. There are biases and contexts that help bootstrap expertise for human faces, a process similar to that of nonperceptual domains (Box 3). Likewise, interest in a particular domain (e.g. bird watching) and contextual cues (e.g. goldfinches almost always appear around a feeder filled with thistle seed) help us acquire expertise for non-face objects.

Explicit feedback

Is it possible that interest/biases and context are sufficient to bootstrap the acquisition of visual expertise? Although the acquisition of face expertise does not appear to require an explicit training signal [43], creating visual expertise in the laboratory requires at least some feedback of observers' recognition performance (Box 3) [44,45,58].

Box 4. Questions for future research

- To what extent are apparent dichotomies in visual recognition explained by two points along a single dimension?
- What types of image and object geometries can be learned by our visual systems at the class level and at more specific levels?
- What aspects of processing change as we become visual experts in a given object domain?
- Is the feedback required to acquire visual expertise completely unsupervised, 'internally' supervised or completely supervised?
- Does learning with one type of feedback lay the groundwork for more efficient learning with additional feedback?
- Is expertise created in the laboratory the same as real-world expertise? If not, how do they differ?
- Are there capacity limits on what we can learn? That is, can we become experts in multiple object domains?
- Does expertise in one domain facilitate the acquisition of expertise in other domains?

The role of explicit feedback in the acquisition of 'real-world' expertise by, for example, birders, dog show judges and car buffs is less clear [47–49] (Box 4). However, similar to the creation of Greeble experts, birders appear to receive feedback from bird guides, car buffs from the manufacturers' nameplates and owner's manuals, and dog show judges from kennel club handbooks. Moreover, face recognition occurs in social contexts that provide much more specific feedback relative to almost any other objects (incorrectly recognizing a goldfinch or a 1965 Mustang has far fewer consequences than failing to recognize one's enemies, friends and offspring).

Concluding remarks

We have argued that all levels of recognition across all object categories can be supported by a single recognition system that is 'tuned' by task, experience and feedback. Although it is tempting to 'divide and conquer' by narrowing the domain of explanation to a subset of object categories or recognition tasks [37], this strategy can lead to false dichotomies that are better explained as points along single dimensions. Such is the case for both the face–object distinction and the specific–basic-level distinction. Experimental and computational findings point towards a single, flexible visual recognition system. For example, Haxby and colleagues [59] report that the representations of faces and several visual categories in the ventral temporal cortex are widely distributed and overlapping, indicating a lack of modularity. We suggest that the case of visual expertise reveals important characteristics about the plasticity of this system.

Acknowledgements

Funding provided by James S. McDonnell Foundation Collaborative Research Award (PEN) and NSF awards BCS-0094491 and DGE-9870676. We wish to thank all of our collaborators in the PEN group for the stimulating discussions that led to many of the ideas presented here.

References

1 Rosch, E. *et al.* (1976) Basic objects in natural categories. *Cogn. Psychol.* 8, 382–439
2 Jolicoeur, P. *et al.* (1984) Pictures and names: making the connection. *Cogn. Psychol.* 16, 243–275
3 Farah, M.J. (1992) Is an object an object an object? Cognitive and neuropsychological investigations of domain-specificity in visual object recognition. *Curr. Dir. Psychol. Sci.* 1, 164–169
4 Farah, M.J. *et al.* (1995) Face perception and within-category discrimination in prosopagnosia. *Neuropsychologia* 33, 661–674
5 Kanwisher, N. *et al.* (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17, 4302–4311
6 McCarthy, G. *et al.* (1997) Face-specific processing in the human fusiform gyrus. *J. Cogn. Neurosci.* 9, 605–610
7 Moscovitch, M. *et al.* (1997) What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *J. Cogn. Neurosci.* 9, 555–604
8 Kanwisher, N. (2000) Domain specificity in face perception. *Nat. Neurosci.* 3, 759–763
9 Goren, C.C. *et al.* (1975) Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics* 56, 544–549
10 Johnson, M.H. and Morton, J. (1991) *Biology and Cognitive Development: The Case of Face Recognition*, Blackwell
11 Yin, R.K. (1969) Looking at upside-down faces. *J. Exp. Psychol.* 81, 141–145
12 Tanaka, J.W. and Farah, M.J. (1993) Parts and wholes in face recognition. *Q. J. Exp. Psychol.* 46, 225–245
13 Perrett, D.I. *et al.* (1982) Visual neurones responsive to faces in the monkey temporal cortex. *Exp. Brain Res.* 47, 329–342
14 Sergent, J. *et al.* (1992) Functional neuroanatomy of face and object processing: a positron emission tomography study. *Brain* 115, 15–36
15 Farah, M.J. (1990) *Visual Agnosia: Disorders of Object Recognition and What They Tell Us About Normal Vision*, MIT Press
16 Gauthier, I. *et al.* (1997) Levels of categorization in visual recognition studied with functional magnetic resonance imaging. *Curr. Biol.* 7, 645–651
17 Gauthier, I. *et al.* (2000) Does visual subordinate-level categorisation engage the functionally defined fusiform face area? *Cogn. Neuropsychol.* 17, 143–163
18 Hamm, J.P. and McMullen, P.A. (1998) Effects of orientation on the identification of rotated objects depend on the level of identity. *J. Exp. Psychol. Hum. Percept. Perform.* 24, 413–426
19 Biederman, I. (1985) Human image understanding: recent research and a theory. *Comput. Vis. Graphics Image Process.* 32, 29–73
20 Jolicoeur, P. (1990) Identification of disoriented objects: a dual-systems theory. *Mind Lang.* 5, 387–410
21 Logothetis, N.K. and Sheinberg, D.L. (1996) Visual object recognition. *Annu. Rev. Neurosci.* 19, 577–621
22 Tarr, M.J. and Bülthoff, H.H. (1998) Image-based object recognition in man, monkey, and machine. *Cognition* 67, 1–20
23 Zhu, S.C. and Yuille, A.L. (1996) FORMS: A flexible object recognition and modeling system. *Int. J. Comput. Vis.* 20, 187–212
24 Koenderink, J.J. and van Doorn, A.J. (1982) The shape of smooth objects and the way contours end. *Perception* 11, 129–137
25 Marr, D. and Nishihara, H.K. (1978) Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. R. Soc. Lond. Ser. B* 200, 269–294
26 Biederman, I. and Gerhardstein, P.C. (1993) Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance. *J. Exp. Psychol. Hum. Percept. Perform.* 19, 1162–1182
27 Edelman, S. (1995) Class similarity and viewpoint invariance in the recognition of 3D objects. *Biol. Cybern.* 72, 207–220
28 Hayward, W.G. and Williams, P. (2000) Viewpoint dependence and object discriminability. *Psychol. Sci.* 11, 7–12
29 Tarr, M.J. (2002) Visual object recognition: can a single mechanism

suffice? *Analytic and Holistic Processes in the Perception of Faces, Objects, and Scenes* (Peterson, M.A. and Rhodes, G., eds), Oxford University Press, (in press)
30 Tarr, M.J. and Bülthoff, H.H. (1995) Is human object recognition better described by geon-structural-descriptions or by multiple-views? *J. Exp. Psychol. Hum. Percept. Perform.* 21, 1494–1505
31 Biederman, I. and Gerhardstein, P.C. (1995) Viewpoint-dependent mechanisms in visual object recognition: Reply to Tarr and Bülthoff (1995). *J. Exp. Psychol. Hum. Percept. Perform.* 21, 1506–1514
32 Tanaka, J.W. and Presnell, L.M. (1999) Color diagnosticity in object recognition. *Percept. Psychophys.* 61, 1140–1153
33 Tarr, M.J. *et al.* (1998) Why the visual system might encode the effects of illumination. *Vision Res.* 38, 2259–2275
34 Perrett, D.I. *et al.* (1998) Evidence accumulation in cell populations responsive to faces: an account of generalisation of recognition without mental transformations. *Cognition* 67, 111–145
35 Tarr, M.J. and Gauthier, I. (2000) FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nat. Neurosci.* 3, 764–769
36 Farah, M.J. *et al.* (2000) Early commitment of neural substrates for face recognition. *Cogn. Neuropsychol.* 17, 117–123
37 Fodor, J.A. (1983) *Modularity of Mind*, MIT Press
38 Pinker, S. (2002) *The Blank Slate: The Modern Denial of Human Nature*, Viking Press
39 Schyns, P. (1998) Diagnostic recognition: Task constraints, object information, and their interactions. *Cognition* 67, 147–179
40 Dixon, M. *et al.* (1997) The interaction of object form and object meaning in the identification performance of a patient with category specific visual agnosia. *Cogn. Neuropsychol.* 114, 1085–1130
41 Rhodes, G. *et al.* (1989) Race Sensitivity in Face Recognition: An Effect of Different Encoding Processes. *Cognition in Individual and Social Contexts* (Bennett, A.F., McConkey, K.M., *et al.* eds), pp. 83–90, Elsevier
42 Carey, S. (1992) Becoming a face expert. *Phil. Trans. R. Soc. London Ser. B* 5, 95–103
43 Gauthier, I. and Tarr, M.J. (1997) Becoming a "Greeble expert": Exploring the face recognition mechanism. *Vision Res.* 37, 1673–1682
44 Gauthier, I. *et al.* (1998) Training Greeble experts: a framework for studying expert object recognition processes. *Vision Res.* 38, 2401–2428
45 Tanaka, J.W. (2001) The entry point of face recognition: evidence for face expertise. *J. Exp. Psychol. Gen.* 130, 534–543
46 Gauthier, I. *et al.* (2000) Expertise for cars and birds recruits brain areas involved in face recognition. *Nat. Neurosci.* 3, 191–197
47 Tanaka, J.W. and Curran, T. (2001) A neural basis for expert object recognition. *Psychol. Sci.* 12, 43–47
48 Diamond, R. and Carey, S. (1986) Why faces are and are not special: An effect of expertise. *J. Exp. Psychol. Gen.* 115, 107–117
49 Gauthier, I. *et al.* (1999) Activation of the middle fusiform "face area" increases with expertise in recognizing novel objects. *Nat. Neurosci.* 2, 568–573
50 Edelman, S. (1993) Representing three-dimensional objects by sets of activities of receptive fields. *Biol. Cybern.* 70, 37–45
51 Quinn, P.C. *et al.* (1993) Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception* 22, 463–475
52 Gold, J. *et al.* (1999) Identification of band-pass filtered letters and faces by human and ideal observers. *Vision Res.* 39, 3537–3560
53 Tanaka, J.W. and Taylor, M. (1991) Object categories and expertise: Is the basic level in the eye of the beholder? *Cogn. Psychol.* 23, 457–482
54 Edelman, S. (1999) *Representation and Recognition in Vision*, MIT Press
55 Hebb, D.O. (1949) *The Organization of Behavior: A Neuropsychological Theory*, John Wiley and Sons
56 Kohonen, T. (1984) *Self-Organization and Associative Memory*, Springer-Verlag
57 Mumford, D. (1992) On the computational architecture of the neocortex: II. The role of the cortico-cortical loop. *Biol. Cybern.* 66, 241–251
58 Gauthier, I. and Tarr, M.J. (2002) Unraveling mechanisms for expert object recognition: bridging brain activity and behavior. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 431–446
59 Haxby, J.V. *et al.* (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293, 2425–2430